Scaling World Models for Agents

From Video Generation to World Model Tutorial @CVPR 2025

Sherry Yang

Nature 2016

ARTICLE

doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

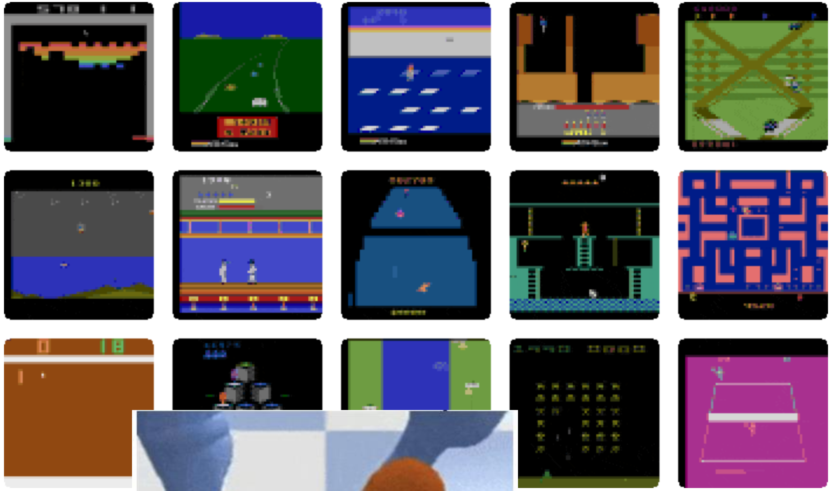
David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

The game of Go has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves. Here we introduce a new approach to computer Go that uses 'value networks' to evaluate board positions and 'policy networks' to select moves. These deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play. Without any lookahead search, the neural networks play Go at the level of state-of-the-art Monte Carlo tree search programs that simulate thousands of random games of self-play. We also introduce a new search algorithm that combines Monte Carlo simulation with value and policy networks. Using this search algorithm, our program AlphaGo achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0. This is the first time that a computer program has defeated a human professional player in the full-sized game of Go, a feat previously thought to be at least a decade away.

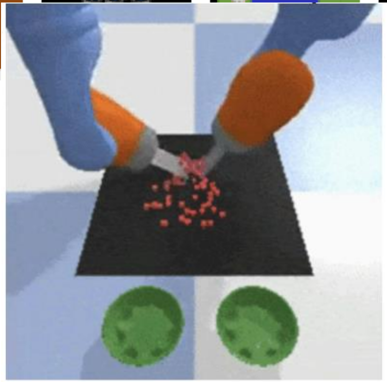
Learning Agents in Simulated Environments

Todorov. MuJoCo. 2012.

Bellemare. Atari. 2012

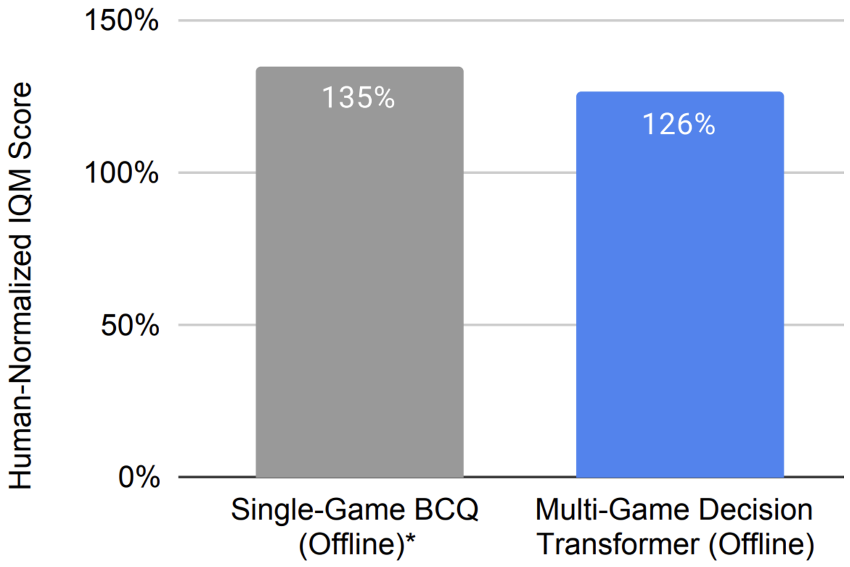


Brockman. 2016

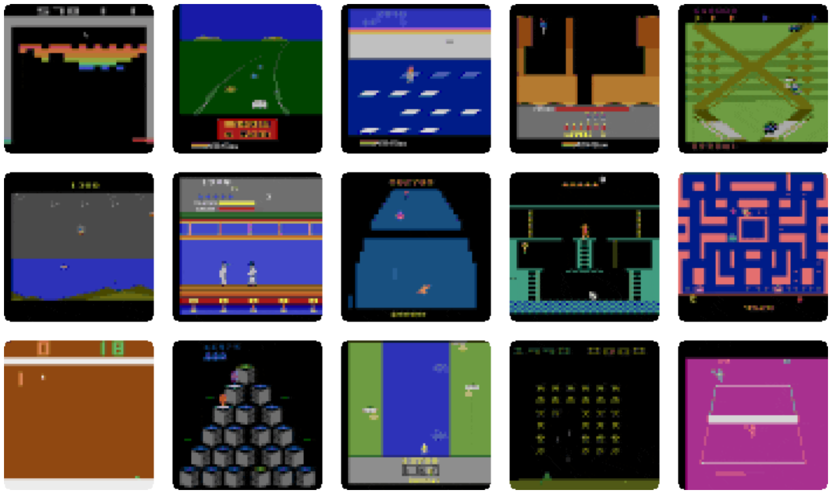


Coumans. Pybullet. 2016

Learning Agents in Multi-Task settings



Bellemare. Atari. 2012



Lee*, Nachum*, **Yang**, Lee, Freeman, Xu, Guadarrama, Fischer, Jang, Michalewski, Mordatch. Multi-Game Decision Transformers. NeurIPS 2022.

Internet Data and Foundation Models

Broad Datasets

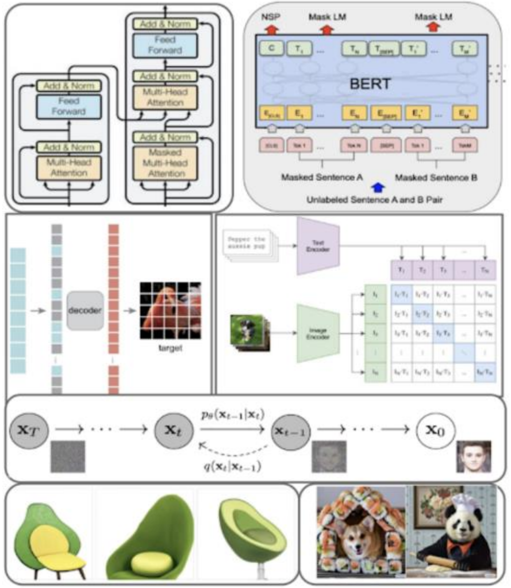
Common Crawl

Wikipedia

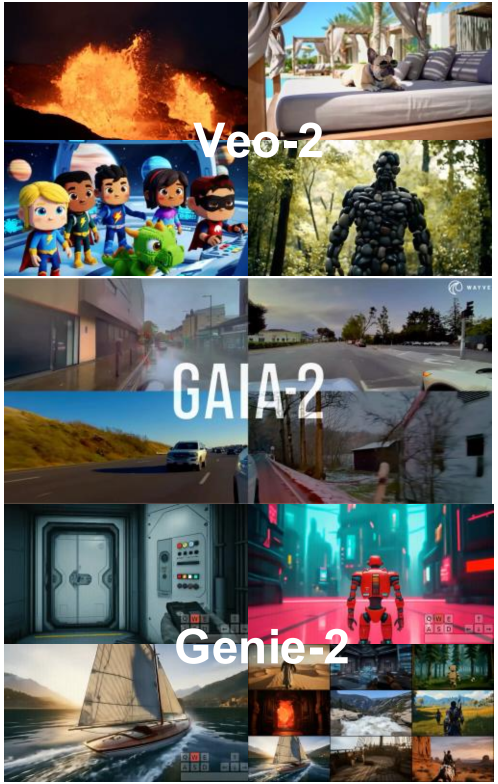
YouTube

Pretrain

Foundation Models



Video Generation



This Talk: Scaling World Models for Agents

Building world models **Scaling data**

- Datasets and modeling
- Action conditioning

Using world models **Scaling computation**

- Long horizon planning
- Evaluating policies
- Training embodied agents

Improving world models **Scaling feedback**

- RL for video generation
- Ground in the physical world through embodied agents

This Talk: Scaling World Models for Agents

Building world models
- Datasets and modeling
- Action conditioning

Using world models
- Long horizon planning
- Evaluating policies
- Training embodied agents

Improving world models
- RL for video generation
- Ground in the physical world through embodied agents

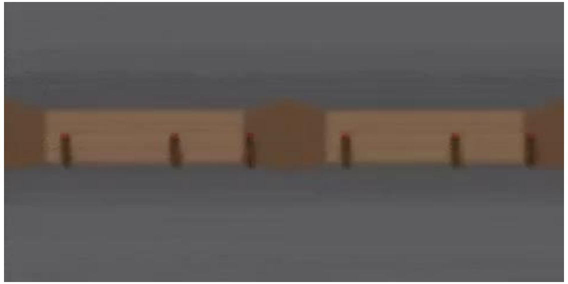
Video Generation as a World Model

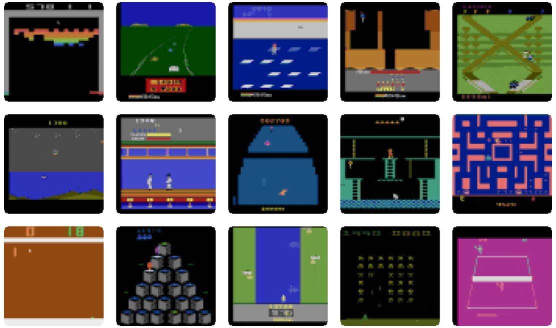
Background: Concept of a world model (dynamics model) existed a while back

Next frames $\mathbf{o}' \sim \hat{T}(\mathbf{o}, \mathbf{a})$ Control actions

Previous frames

Ground truth Generated





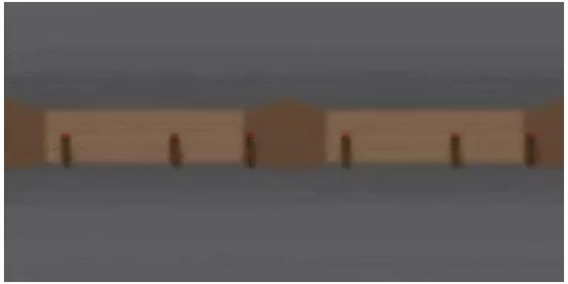
Ha and Schmidhuber. Recurrent World Models Facilitate Policy Evolution. NeurIPS 2018.
Hafner, Lillicrap, Ba, Norouzi. Dream to Control: Learning Behaviors by Latent Imagination. ICLR 2020.

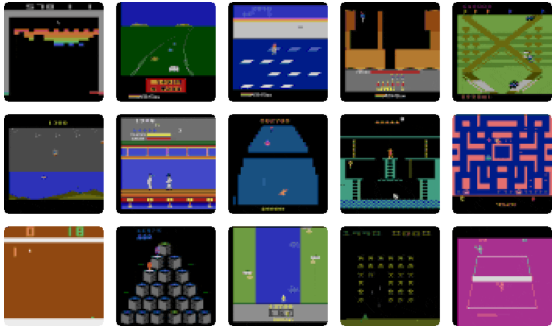
Video Generation as a World Model

Background: Concept of a world model (dynamics model) existed a while back

Question: What is different now?

Ground truth Generated





Ha and Schmidhuber. Recurrent World Models Facilitate Policy Evolution. NeurIPS 2018.
Hafner, Lillicrap, Ba, Norouzi. Dream to Control: Learning Behaviors by Latent Imagination. ICLR 2020.

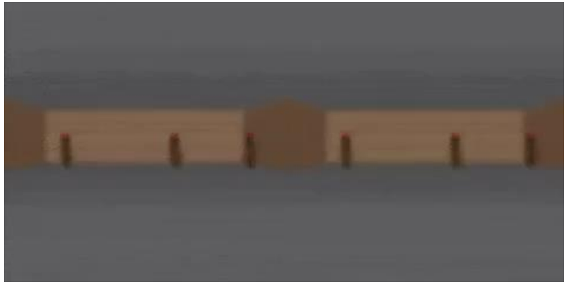
Video Generation as a World Model

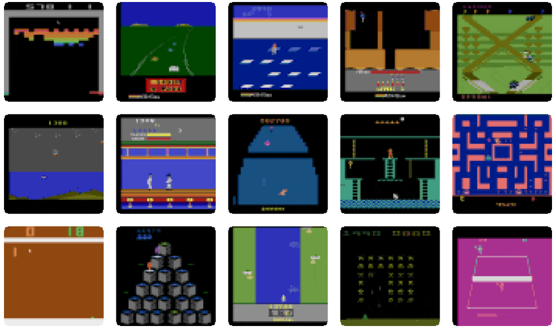
Background: Concept of a world model (dynamics model) existed a while back

Question: What is different now?

- Internet-scale dataset **Realistic world simulators**

Ground truth Generated





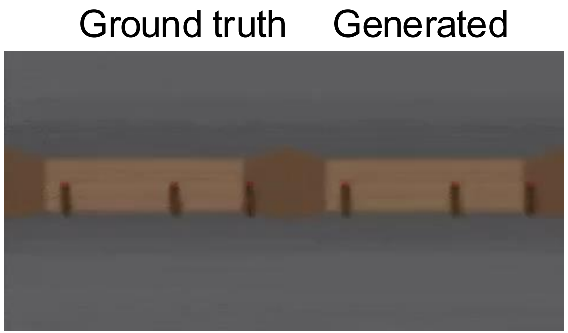
[Ha and Schmidhuber. Recurrent World Models Facilitate Policy Evolution. NeurIPS 2018.]
[Hafner, Lillicrap, Ba, Norouzi. Dream to Control: Learning Behaviors by Latent Imagination. ICLR 2020.]

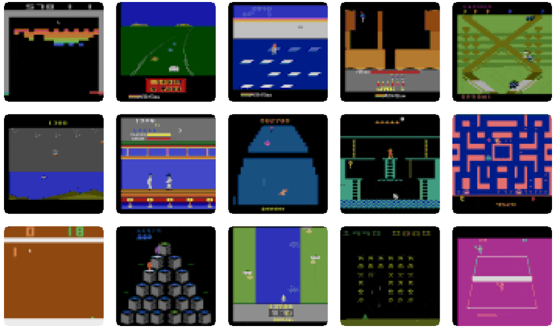
Video Generation as a World Model

Background: Concept of a world model (dynamics model) existed a while back

Question: What is different now?

- Internet-scale dataset **Realistic world simulators**
- Scalable video generation architectures **Single world model across environments**





Ha and Schmidhuber. Recurrent World Models Facilitate Policy Evolution. NeurIPS 2018.
Hafner, Lillicrap, Ba, Norouzi. Dream to Control: Learning Behaviors by Latent Imagination. ICLR 2020.

Internet-Scale Dataset for World Modeling

Any time-aligned video-"action" data

Text-video pairs:

A person cutting the pepper with a knife

Time

Yang, Walker, Parker-Holder, Du, Bruuce, Barreto, Abbeel, Schuurmans. Video as the New Language for Real-World Decision Making. ICML 2024.

Internet-Scale Dataset for World Modeling

Any time-aligned video-"action" data

Camera control:

Turn 360 degrees clockwise

Time

Yang, Walker, Parker-Holder, Du, Bruuce, Barreto, Abbeel, Schuurmans. Video as the New Language for Real-World Decision Making. ICML 2024.

Internet-Scale Dataset for World Modeling

Any time-aligned video-"action" data

Robot control:

$\Delta x, \Delta y$ $\Delta x, \Delta y$

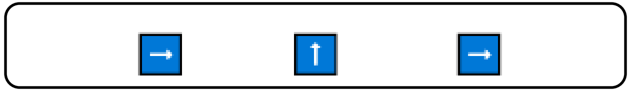
Time

Yang, Walker, Parker-Holder, Du, Bruuce, Barreto, Abbeel, Schuurmans. Video as the New Language for Real-World Decision Making. ICML 2024.

Internet-Scale Dataset for World Modeling

Any time-aligned video-"action" data

Keyboard control:



Time

[**Yang**, Walker, Parker-Holder, Du, Bruuce, Barreto, Abbeel, Schuurmans. Video as the New Language for Real-World Decision Making. ICML 2024.]

Internet-Scale Dataset for World Modeling

Any time-aligned video-"action" data

A person cutting the pepper with a knife

$\Delta x, \Delta y$ $\Delta x, \Delta y$

Turn 360 degrees clockwise

Training data (21M video-"action" pairs)

Yang, Du, Ghasemipour, Tompson, Kaelbling, Schuurmans, Abbeel. Learning Interactive Real-World Simulators. ICLR 2024.

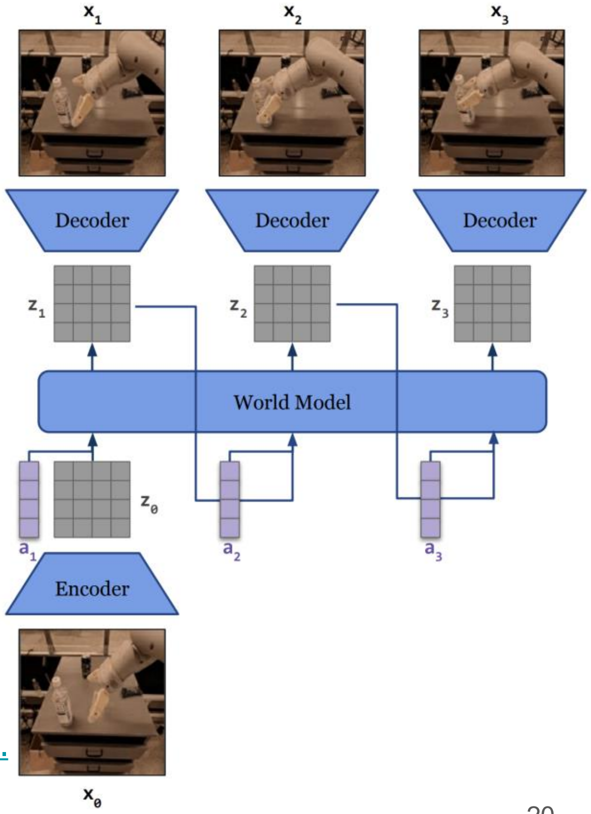
Scalable Video Generation Architectures

Video diffusion models: 3D UNet (DiT/latent diffusion)

Classifier-free guidance: Text conditioning

Model cascade: Temporal and spatial super-resolution

Image conditioning: Block-wise autoregressive rollouts



Ho, et al. Video Diffusion Models. ICLR 2022.

Peebles and Xie. Scalable Diffusion Models with Transformers. ICCV 2023.

Ho and Salimans. Classifier-Free Diffusion Guidance. NeurIPS 2021.

Ho*, Saharia*, et al. Cascaded Diffusion Models for High Fidelity Image Generation. JMLR 2022.

Ho, et al. Imagen Video: High Definition Video Generation with Diffusion Models. arXiv 2022.

Chen, et al. Next-token Prediction Meets Full-Sequence Diffusion. NeurIPS 2024.

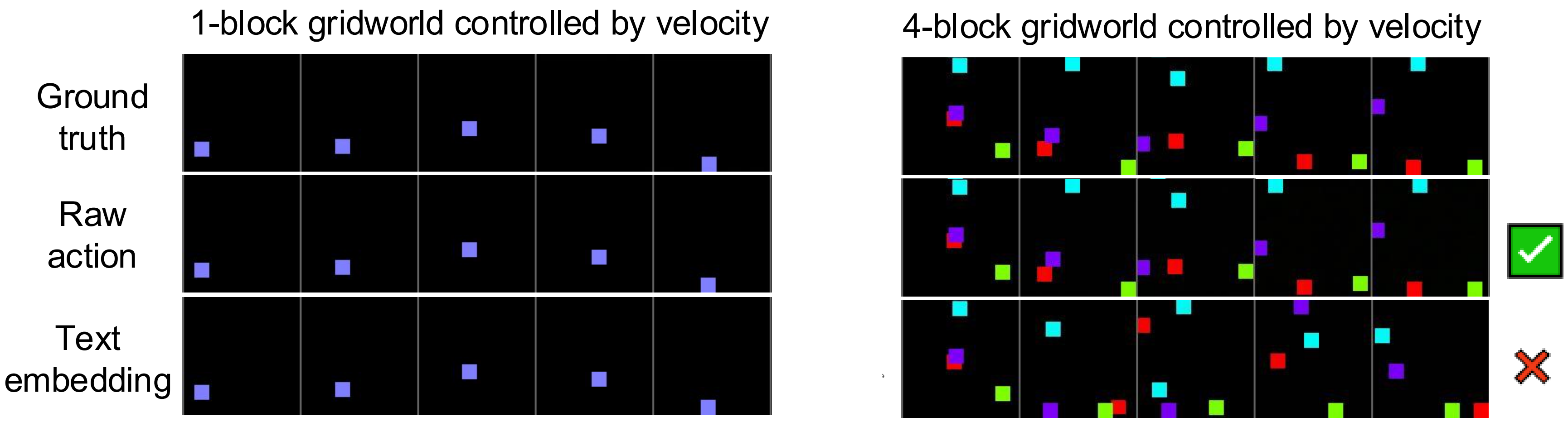
Action Conditioning

Question: How to represent continuous control actions?

Action Conditioning

Question: How to represent continuous control actions?

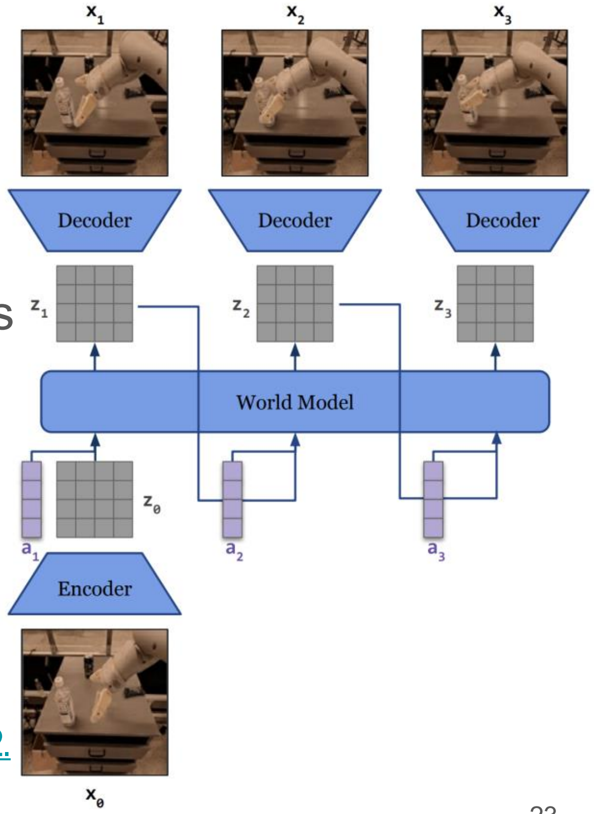
- Use text embeddings (LLM, CLIP, T5), discretization
- Use the original continuous vector



[Quevedo, Liang, **Yang**. Evaluating Robot Policies in a World Model. arXiv 2025.]

Scalable Video Generation Architectures

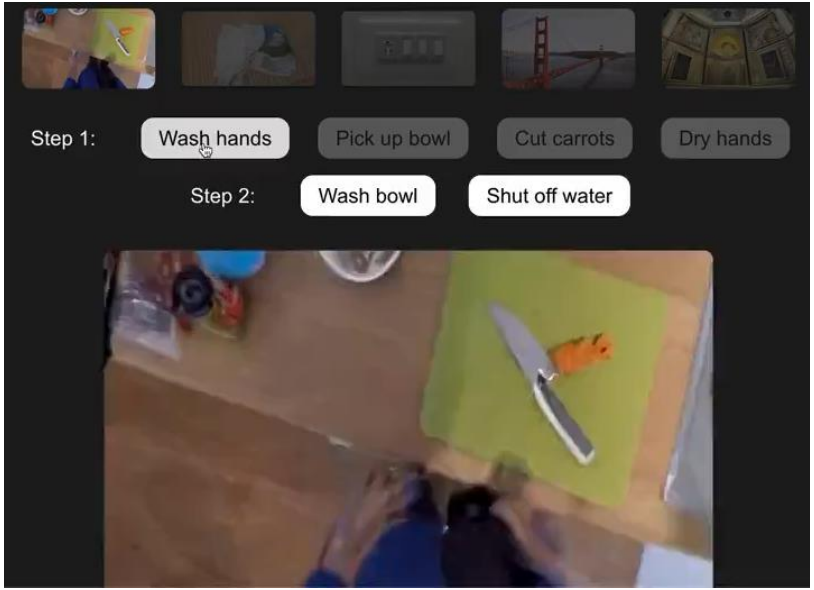
Video diffusion models: 3D UNet (DiT/latent diffusion)
Classifier-free guidance: Text conditioning
Model cascade: Temporal and spatial super-resolution
Image conditioning: Block-wise autoregressive rollouts
Action conditioning: Linear projection of raw continuous vectors



Ho, et al. Video Diffusion Models. ICLR 2022.
Peebles and Xie. Scalable Diffusion Models with Transformers. ICCV 2023.
Ho and Salimans. Classifier-Free Diffusion Guidance. NeurIPS 2021.
Ho*, Saharia*, et al. Cascaded Diffusion Models for High Fidelity Image Generation. JMLR 2022.
Ho, et al. Imagen Video: High Definition Video Generation with Diffusion Models. arXiv 2022.
Chen, et al. Next-token Prediction Meets Full-Sequence Diffusion. NeurIPS 2024.

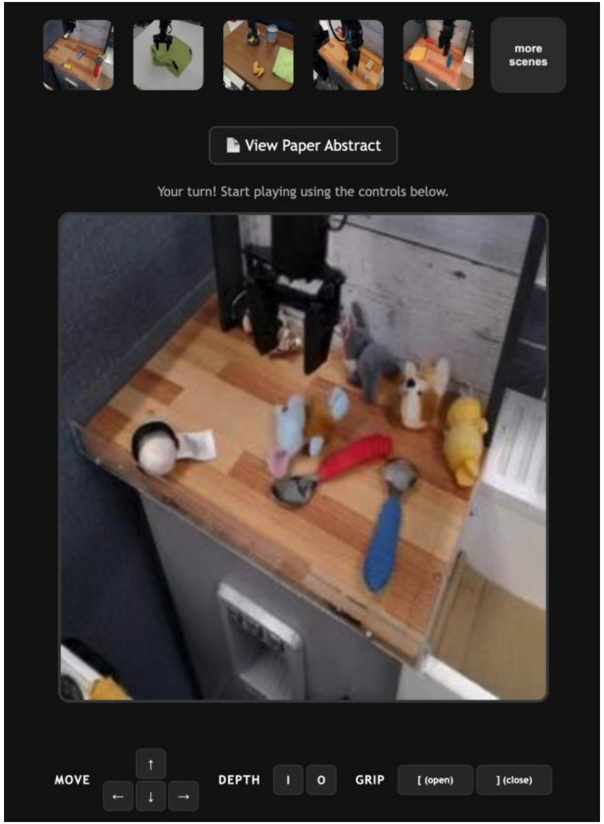
Examples – Try Yourself!

universal-simulator.github.io



5B, 512 TPUs, 20 days

world-model-eval.github.io



600M,
2 A100,
5 days

Yang, Du, Ghasemipour, Tompson, Kaelbling, Schuurmans, Abbeel. Learning Interactive Real-World Simulators. ICLR 2024.
Quevedo, Liang, **Yang**. Evaluating Robot Policies in a World Model. arXiv 2025.

This Talk: Scaling World Models for Agents

Building world models

Scaling data: time-aligned video-"action"

- Datasets and modeling
- Action conditioning

Using world models

- Long horizon planning
- Evaluating policies
- Training embodied agents

Improving world models

- RL for video generation
- Ground in the physical world through embodied agents

This Talk: Scaling World Models for Agents

Building world models

Scaling data: time-aligned video-"action"

- Datasets and modeling
- Action conditioning

Using world models

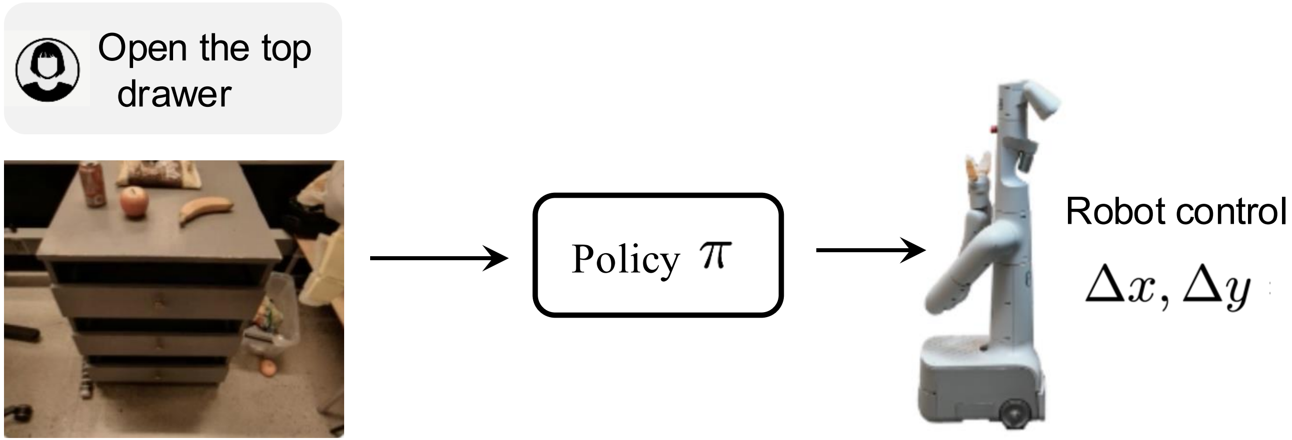
- Long horizon planning
- Evaluating policies
- Training embodied agents

Improving world models

- RL for video generation
- Ground in the physical world through embodied agents

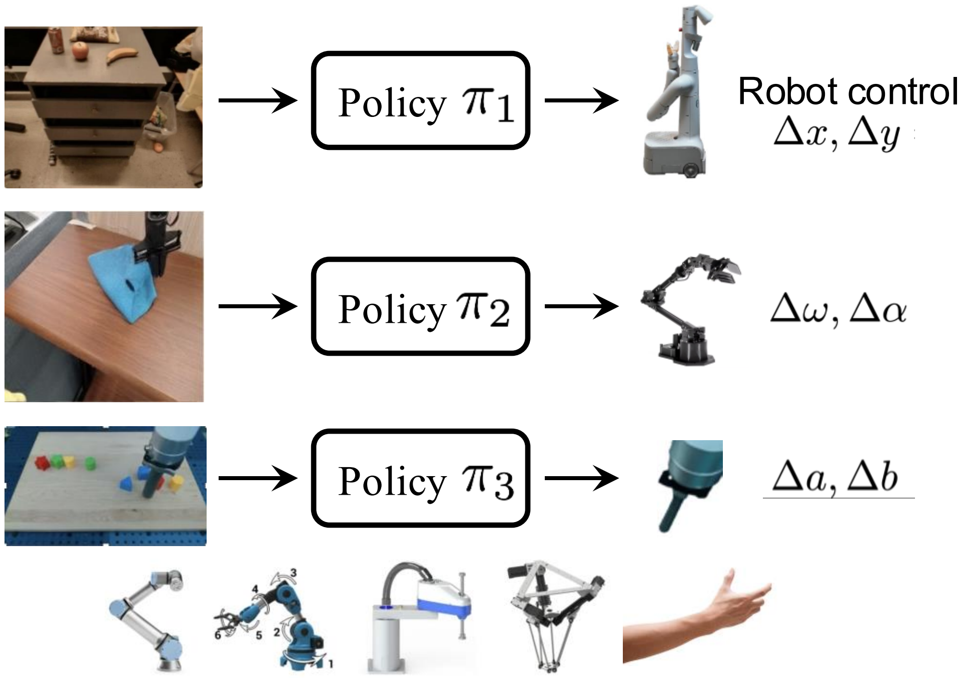
Planning in a World Model

Problem: Learning control policies mapping from observation to action



Planning in a World Model

Prior approach: Learning one policy for each environment and each robot



No knowledge sharing

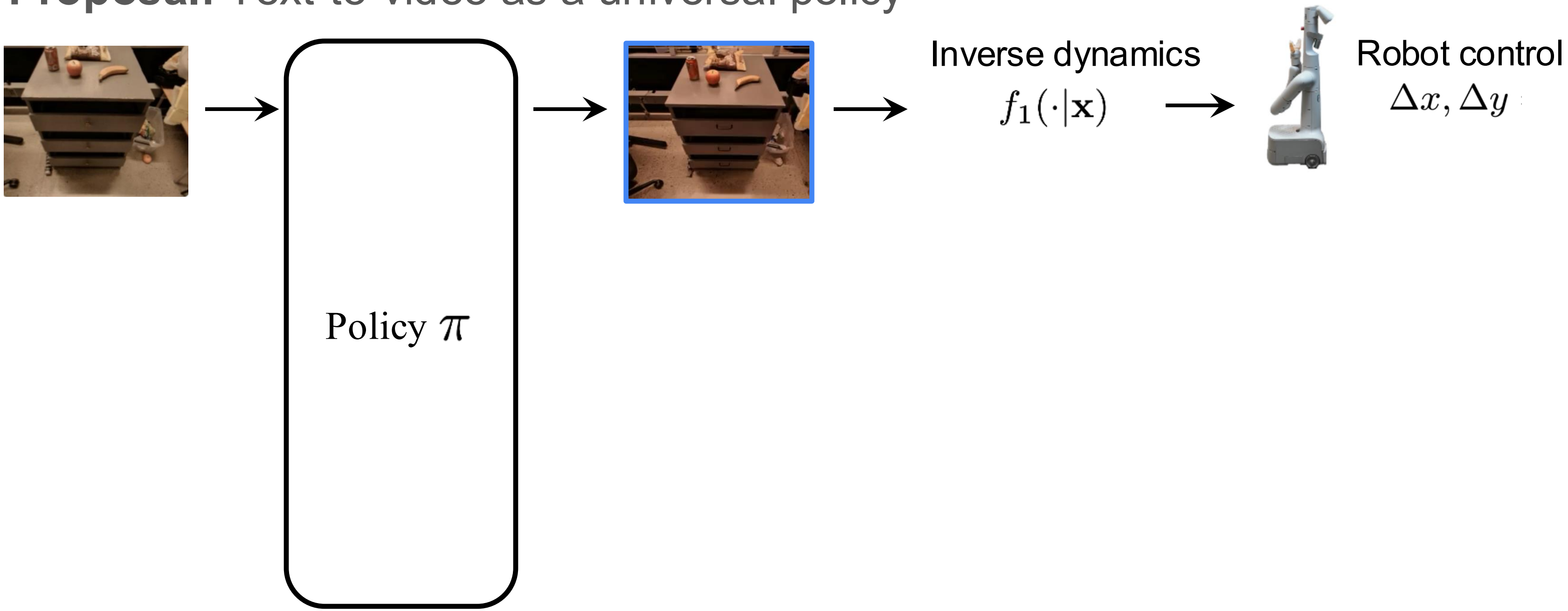
Planning in a World Model

Proposal: Text-to-video as a universal policy

Du*, **Yang***, et al. Learning Universal Policies via Text-Guided Video Generation. NeurIPS 2023

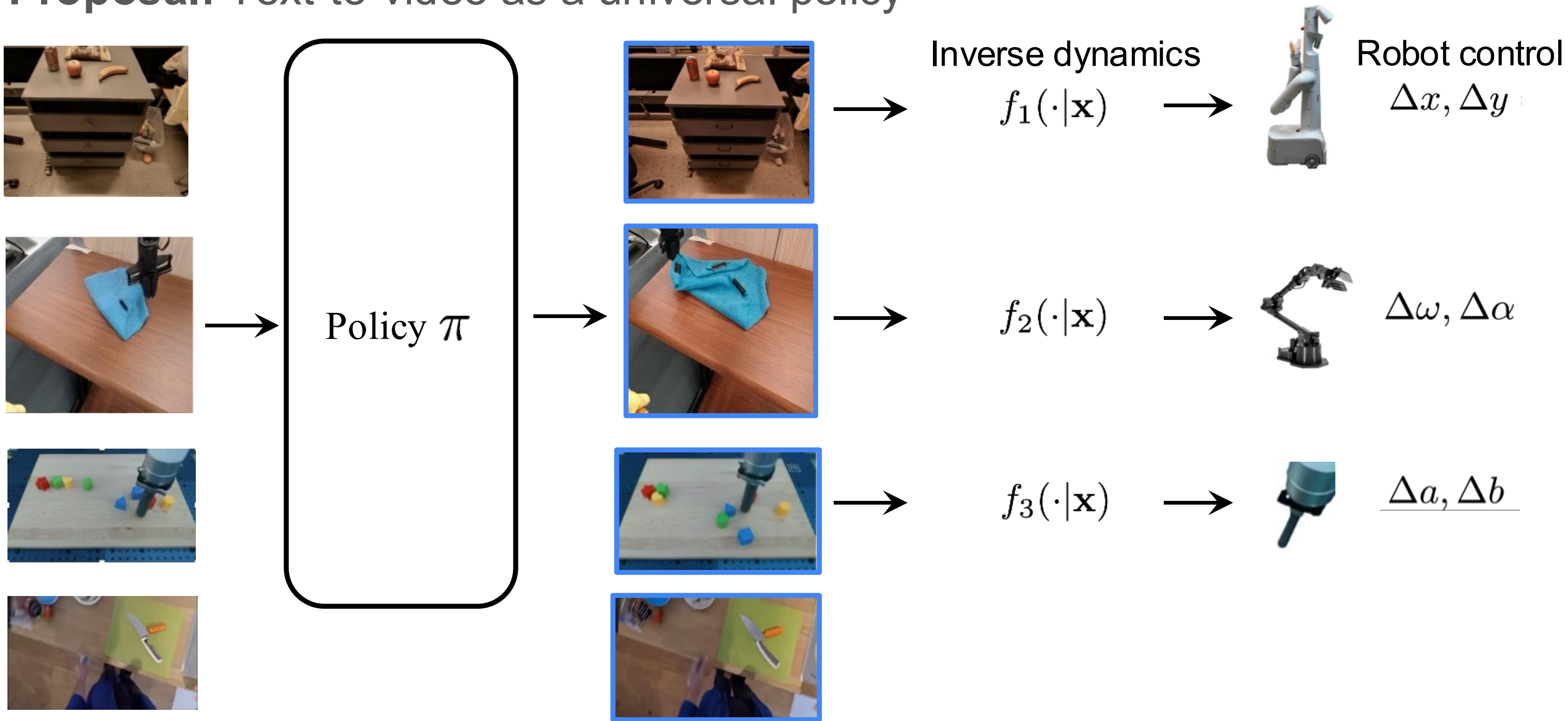
Planning in a World Model

Proposal: Text-to-video as a universal policy



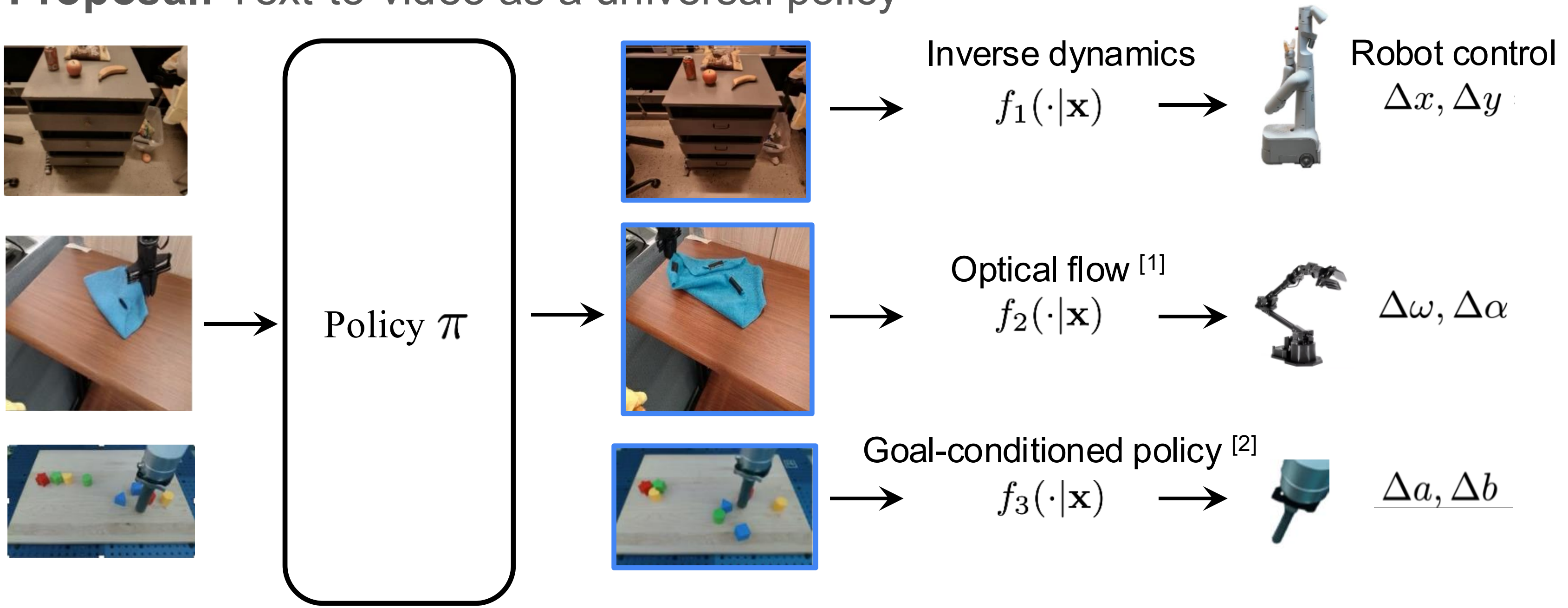
Planning in a World Model

Proposal: Text-to-video as a universal policy



Planning in a World Model

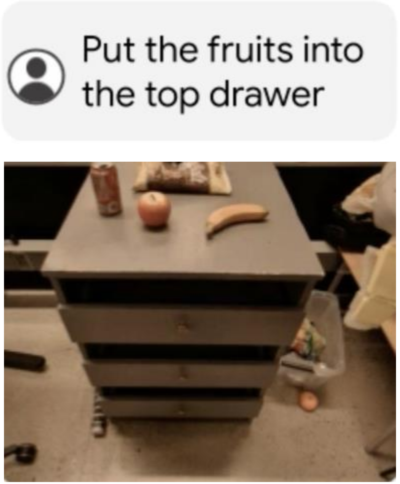
Proposal: Text-to-video as a universal policy



[1] Ko, et al. Learning to Act from Actionless Videos through Dense Correspondences. ICLR 2024.
[2] Black, et al. Zero-Shot Robotic Manipulation with Pretrained Image-Editing Diffusion Models. ICLR 2024.

Long Horizon Planning in a World Model

Challenge: Hard to generate a complex step-by-step video in one go

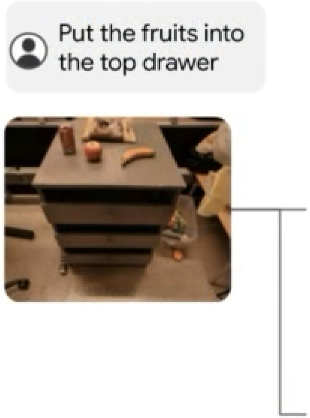


Generate plans one step at a time

1) Open top drawer
2) Put banana in the top drawer
3) Put apple in the top drawer
4) Close top drawer

Long Horizon Planning in a World Model

Planning in the video and language space



Du, **Yang**, Florence, Xia, Wahid, Ichter, Sermanet, Yu, Abbeel, Tenenbaum, Kaelbling, Zeng, Tompson. Video Language Planning. ICLR 2024.

Long Horizon Planning in a World Model

Put all fruits in the top drawer

Generated video

Real-world execution

$\longrightarrow f(\cdot|\mathbf{x})$

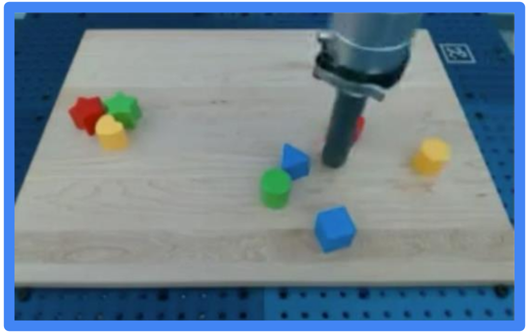
$\longrightarrow$

Du, **Yang**, Florence, Xia, Wahid, Ichter, Sermanet, Yu, Abbeel, Tenenbaum, Kaelbling, Zeng, Tompson. Video Language Planning. ICLR 2024.

Long Horizon Planning in a World Model

Make a line

Generated video



Real-world execution

1) Move the red circle to the left of the yellow hexagon
2) Move the green circle closer to the red star
3) Move the blue triangle to the top left of the red circle
4) Move the blue cube to the left of the blue triangle
5) Move the green circle to the center
6) Push the green circle towards the yellow heart
7) Move the blue triangle to the right of the green circle
8) Slide the blue cube towards the blue triangle
9) Push the red circle closer to the blue cube
10) Move the yellow hexagon closer to the red circle

Beams	Language Branch	Video Branch	Line Performance
1	1	1	4%
1	1	4	10%
1	4	4	22%
2	4	4	**56%**

Evaluating Policies in a World Model

How good is a policy π ?

Evaluating Policies in a World Model

How good is a policy π ?

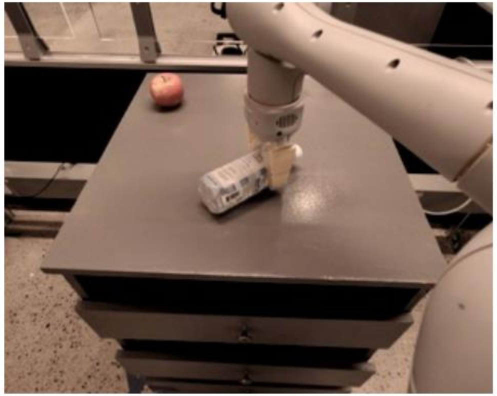
- Run on the real robot “My lab has 5 PhD students and 1 robot” “and the robot broke”

Evaluating Policies in a World Model

How good is a policy π ?

- Run on the real robot "My lab has 5 PhD students and 1 robot" "and the robot broke"
- Run in software simulator Poor correlation between simulated and real-world outcomes

Real world



Software simulator

Evaluating Policies in a World Model

How good is a world model for policy evaluation?

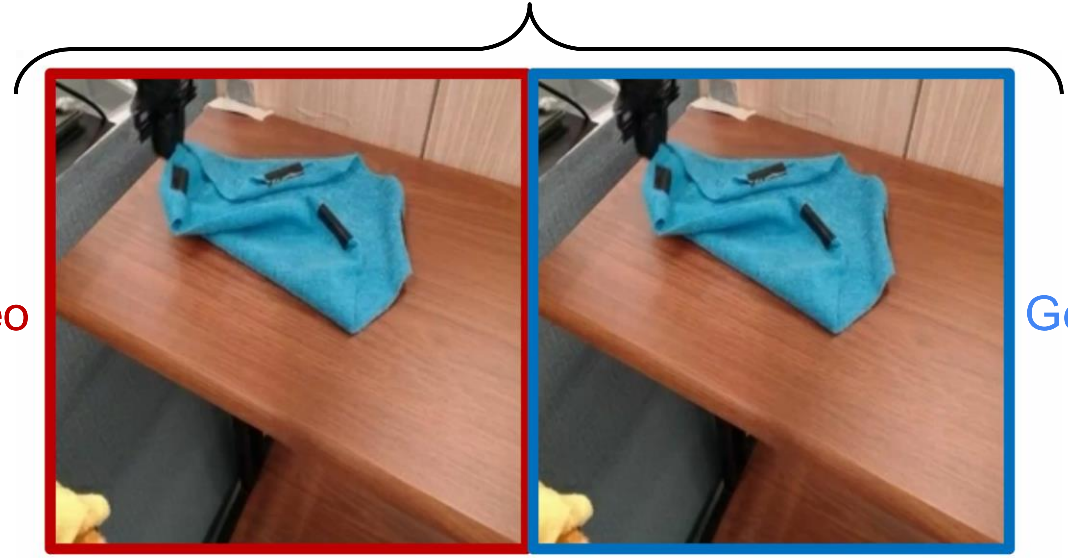
Quevedo, Liang, **Yang**. Evaluating Robot Policies in a World Model. 2025.

Evaluating Policies in a World Model

How good is a world model for policy evaluation?

Same sequence of actions: $\Delta x, \Delta y, \Delta \omega, \Delta \alpha, \Delta a, \Delta b$

Real video

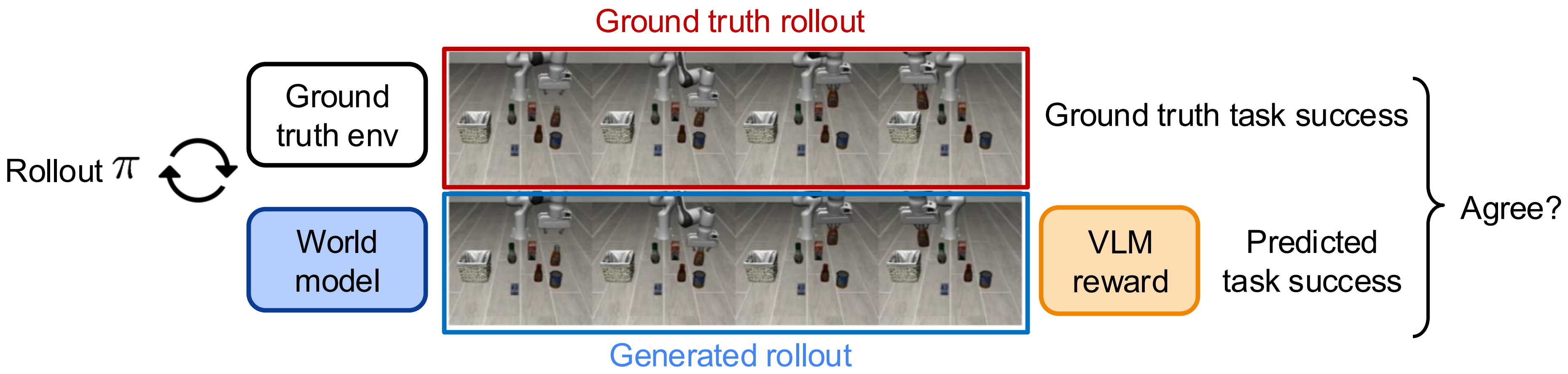


Generated video

[Quevedo, Liang, **Yang**. Evaluating Robot Policies in a World Model. 2025.]
[Li, et al. WorldEval: World Model as Real-World Robot Policies Evaluator. 2025]

Evaluating Policies in a World Model

How good is a world model for policy evaluation?

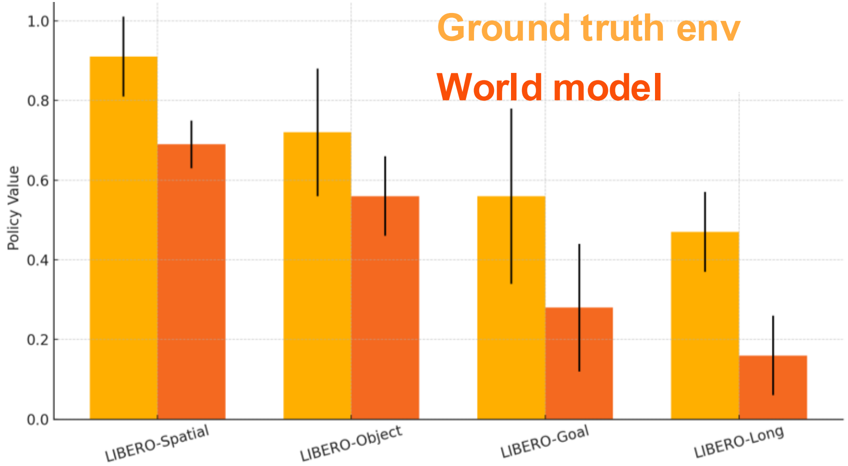


[Quevedo, Liang, **Yang**. Evaluating Robot Policies in a World Model. 2025.]
[Li, et al. WorldEval: World Model as Real-World Robot Policies Evaluator. 2025]

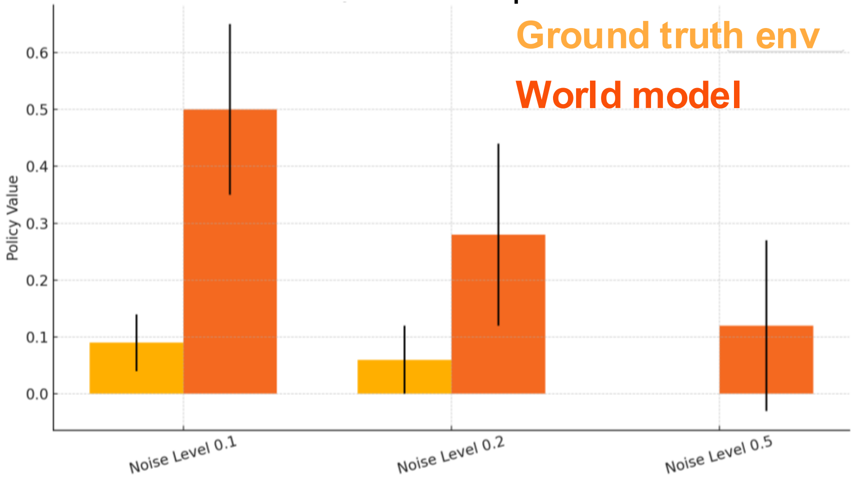
Evaluating Policies in a World Model

How good is a world model for policy evaluation?

In-distribution (data collection policy)
False negative



Out-of-distribution (noisy policy)
False positive



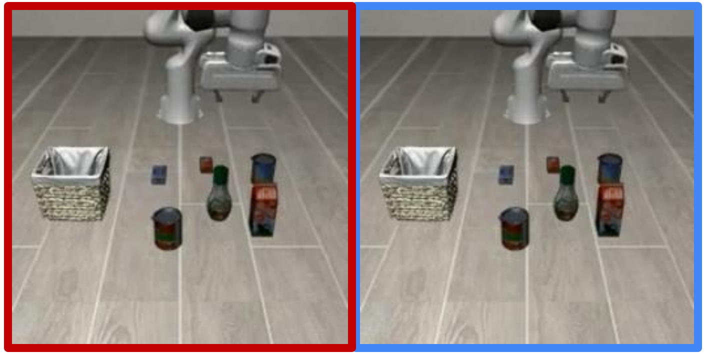
[Quevedo, Liang, **Yang**. Evaluating Robot Policies in a World Model. 2025.]()

Evaluating Policies in a World Model

How good is a world model for policy evaluation?

In-distribution (data collection policy)
False negative

Ground truth　Generated



Out-of-distribution (noisy policy)
False positive

Ground truth　Generated

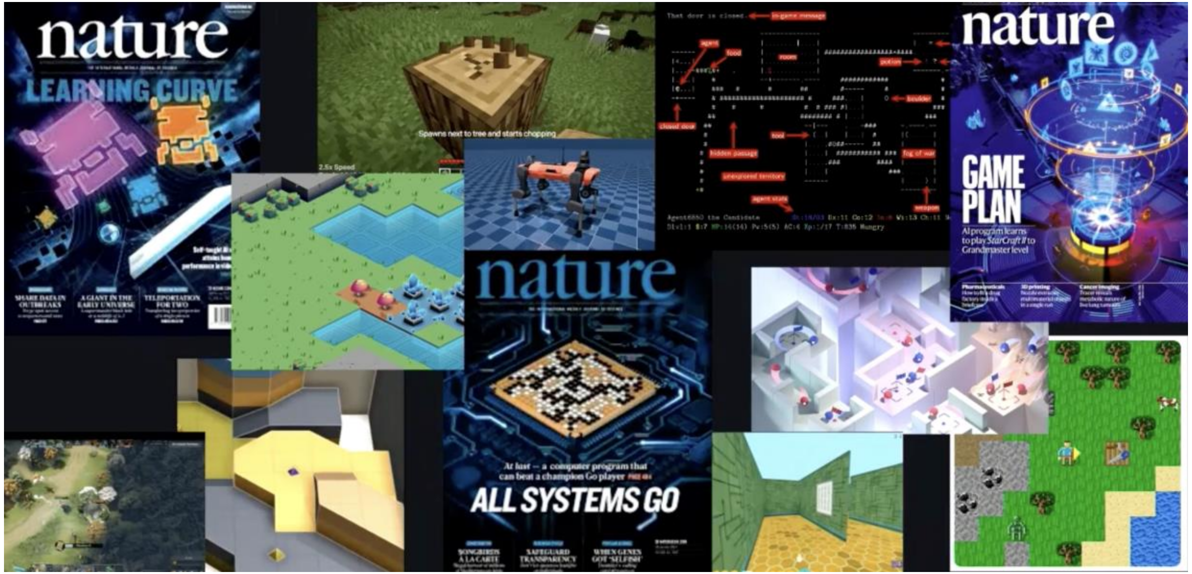
[Quevedo, Liang, **Yang**. Evaluating Robot Policies in a World Model. 2025.]

Improve Policies in a World Model

Improve Policies in a World Model

We know how to improve policies in a simulator (Go, Atari, Starcraft)

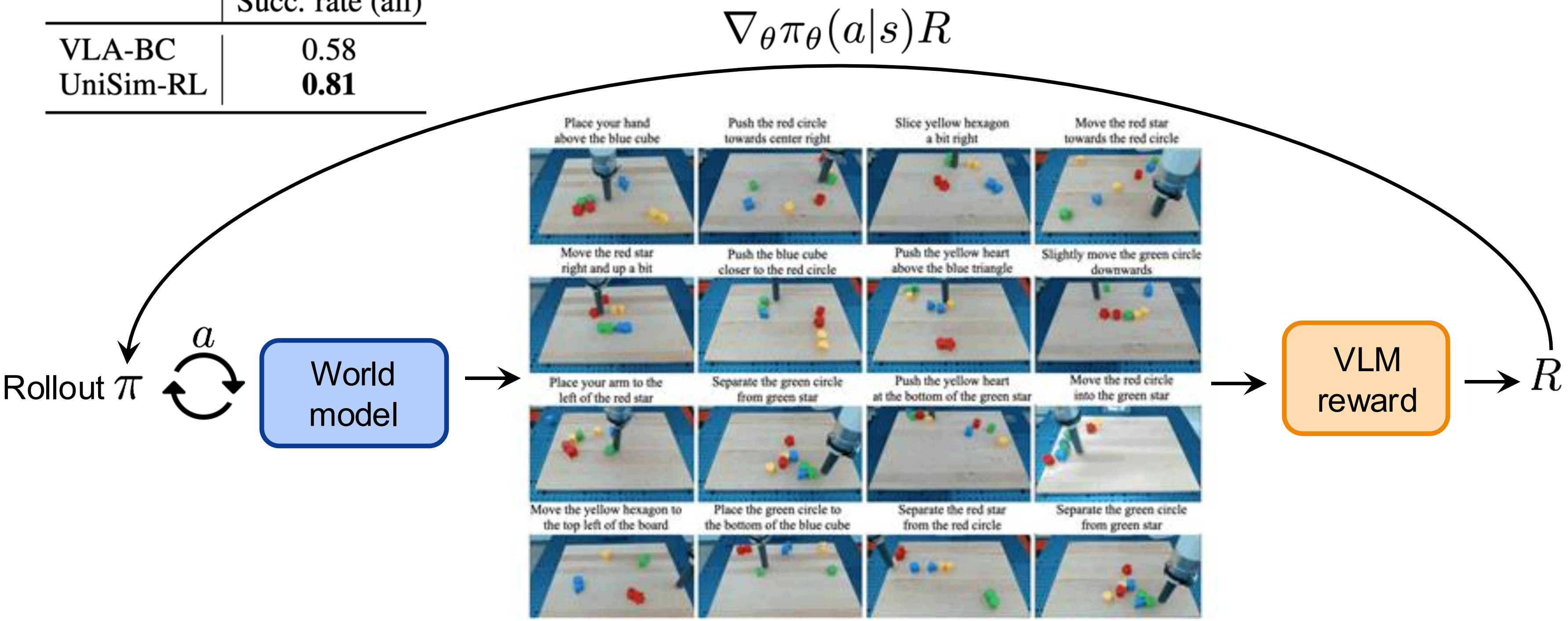
to achieve super-human performance



Improve Policies in a World Model

Running RL (policy gradient) using rollouts from the world model

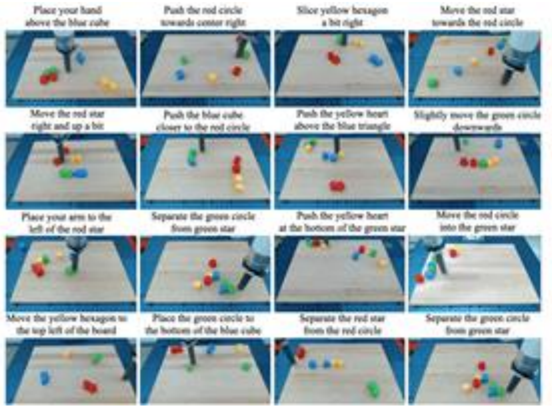
	Succ. rate (all)
VLA-BC	0.58
UniSim-RL	**0.81**



Yang, Du, Ghasemipour, Tompson, Kaelbling, Schuurmans, Abbeel. Learning Interactive Real-World Simulators. ICLR 2024.

Improve Policies in a World Model

Train in world model



Test in real world

Push the red star towards the blue cube

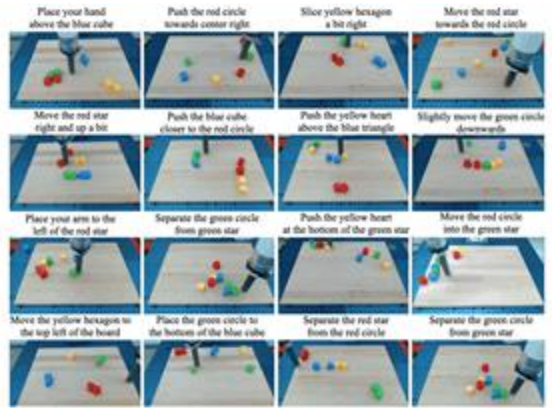
Yang, Du, Ghasemipour, Tompson, Kaelbling, Schuurmans, Abbeel. Learning Interactive Real-World Simulators. ICLR 2024.

Improve Policies in a World Model

Algorithm itself is similar to model-based RL

Difference: Real-world tasks (beyond games)

Train in world model



Test in real world

Push the red star towards the blue cube

Kaelbling, Littman, Moore. Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research 1996.
Ha and Schmidhuber. Recurrent World Models Facilitate Policy Evolution. NeurIPS 2018.
Hafner, Lillicrap, Ba, Norouzi. Dream to Control: Learning Behaviors by Latent Imagination. ICLR 2020.
Kaiser, et al. Model-Based Reinforcement Learning for Atari. ICLR 2020.

This Talk: Scaling World Models with Agents

Building world models **Scaling data: time-aligned video-"action"**

- Datasets and modeling
- Action conditioning

Using world models **Scaling computation: search, planning, rolling out in a real-world simulator**

- Long horizon planning
- Evaluating policies
- Training embodied agents

Improving world models

- RL for video generation
- Ground in the physical world through embodied agents

This Talk: Scaling World Models with Agents

Building world models **Scaling data: time-aligned video-"action"**

- Datasets and modeling
- Action conditioning

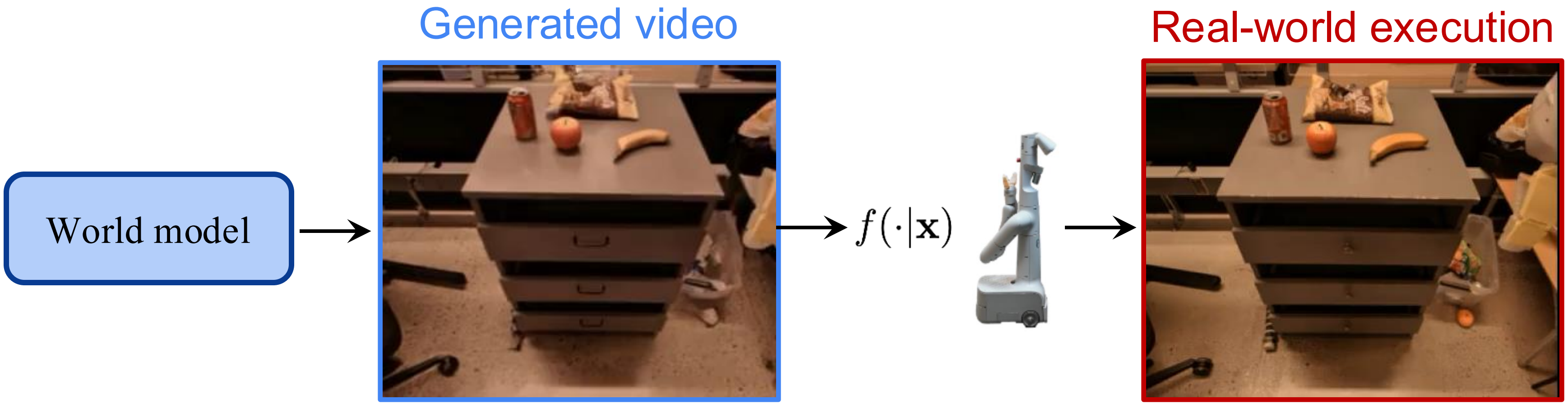
Using world models **Scaling computation: search, planning, rolling out in a real-world simulator**

- Long horizon planning
- Evaluating policies
- Training embodied agents

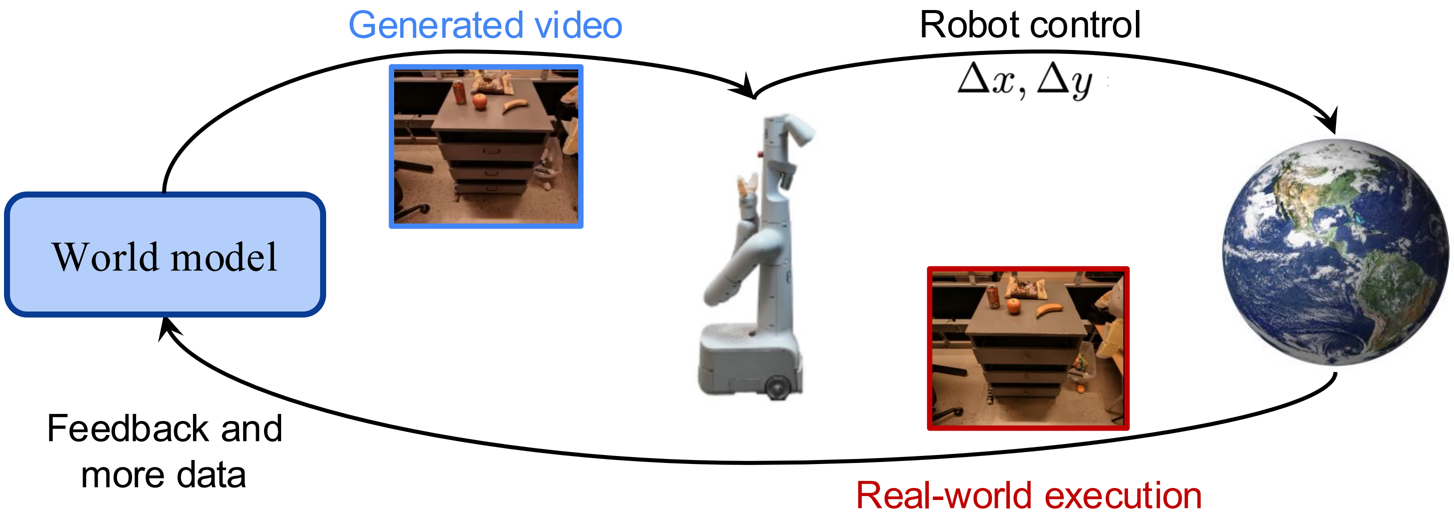
Improving world models

- RL for video generation
- Ground in the physical world through embodied agents

Planning in a World Model

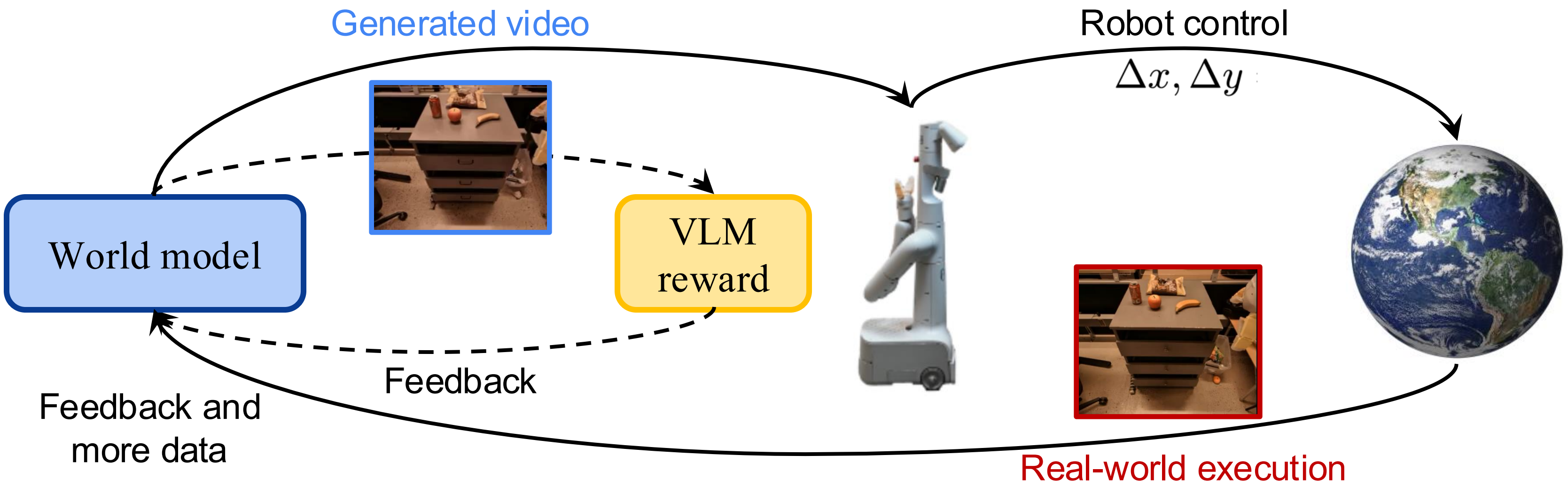


Improving a World Model with Feedback



Soni*, Venkataraman*, Chandra*, Fischmeister, Liang, Dai, **Yang**. VideoAgent: Self-Improving Video Generation. 2025.

Improving a World Model with Feedback



Soni*, Venkataraman*, Chandra*, Fischmeister, Liang, Dai, **Yang**. VideoAgent: Self-Improving Video Generation. 2025.

Improving a World Model with Feedback

With VLM feedback:

- Training to self-correct
- Reinforcement learning for video generation (e.g., DPO)

Kumar, et al. Training Language Models to Self-Correct via Reinforcement Learning. ICLR 2025.
Chen, et al. Teaching Large Language Models to Self-Debug. ICLR 2024.
Soni*, Venkataraman*, Chandra*, Fischmeister, Liang, Dai, **Yang**. VideoAgent: Self-Improving Video Generation. 2025.
Furuta, Zen, Schuurmans, Faust, Matsuo, Liang, **Yang**. Improving Text-to-Video Generation with AI Feedback. 2025

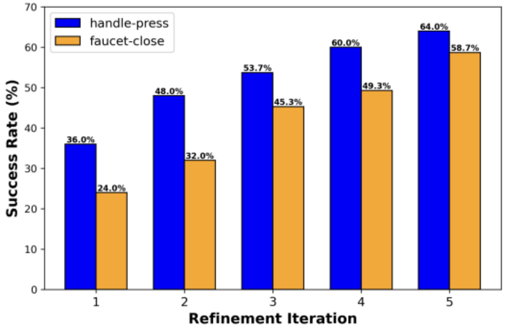
Improving a World Model with Feedback

With VLM feedback:

- Training to self-correct
- Reinforcement learning for video generation (e.g., DPO)

With execution feedback:

- Iterative learning and data generation (e.g., DAgger, STaR)



Ross, Gordon, Bagnell. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. AISTATS 2011.
Zelikman, Wu, Mu, Goodman. STaR: Bootstrapping Reasoning With Reasoning. NeurIPS 2022.
Soni*, Venkataraman*, Chandra*, Fischmeister, Liang, Dai, **Yang**. VideoAgent: Self-Improving Video Generation. 2025.

This Talk: Scaling World Models for Agents

Building world models **Scaling data: time-aligned video-"action"**

- Dataset and modeling
- Action conditioning

Using world models **Scaling computation: search, planning, rolling out in a real-world simulator**

- Long horizon planning
- Evaluating policies
- Training embodied agents

Improving world models **Scaling feedback: AI and execution**

- RL for video generation
- Ground in the physical world through embodied agents

Final Remarks

Dream: **Universal** environment for agents

- Through computer vision
- Promise of generalization

Key: World models

- Useful signals from broad data
- Understand counterfactuals, simulate different outcomes
- Do long horizon **planning** (at different abstraction levels with language and video)

Think about safety

- Any video you see on a computer can be hijacked by a world mode
- Something to step up if we are going to use a world model to train general purpose agents

Scaling World Models for Agents

Building world models **Scaling data: time-aligned video-"action"**

- Dataset and modeling
- Action conditioning

Using world models **Scaling computation: search, planning, rolling out in a real-world simulator**

- Long horizon planning
- Evaluating policies
- Training embodied agents

Improving world models **Scaling feedback: AI and execution**

- RL for video generation
- Ground in the physical world through embodied agents